EXECUTIVE SUMMARY

Introduction

This report describes the conclusions and recommendations of the Expert Panel ("Panel") regarding the validation status of four *in vitro* ocular toxicity test methods: the Isolated Rabbit Eye (IRE), the Isolated Chicken Eye (ICE), the Bovine Corneal Opacity and Permeability (BCOP), and the Hen's Egg Test - Chorioallantoic Membrane (HET-CAM) assays. Those areas of each background review document (BRD) not mentioned in this report were considered adequate and acceptably accurate by the Panel.

The Isolated Rabbit Eye Test Method

The Panel concluded that the IRE BRD proposed version of the IRE test method appears to be capable of identifying ocular corrosives/severe irritants in a tiered-testing strategy with the caveat that the accuracy of this test method be corroborated using a larger number of substances and that reliability analyses be conducted when additional data become available. This recommendation was based on the relatively small number of substances (n=36) tested using the proposed IRE test method version and because only one laboratory (SafePharm, Derby, United Kingdom) had experience using this test method protocol. The Panel agreed that the recommended standardized protocol described in the IRE BRD, which included fluorescein penetration and evaluation of epithelial integrity as endpoints, was appropriate and significantly improved accuracy when compared to other versions of the IRE test method.

With respect to IRE optimization and validation, the Panel recommended that additional data be requested from users of this test method and that analyses of additional data be conducted. The Panel also suggested, that as the IRE test method had a relatively high false positive rate of 33% (with a false negative rate of 0%), optimization of the decision criteria to minimize the false positive rate without appreciably increasing the false negative rate is needed. This may best be accomplished using statistical methods (e.g., discriminant analysis) to improve the decision criteria for the IRE. The Panel noted that any further optimization or validation should be conducted using existing data. Additional animal studies would only be conducted if important data gaps were identified and such studies would be carefully designed to maximize the amount of pathophysiological information obtained (e.g., wound healing). A minority opinion of one Panel member stated that no additional animals should be used for this purpose. The Panel also recommended that a high quality database of *in vivo* and *in vitro* data of reference substances be established from existing literature and new data.

The Panel proposed several modifications to the recommended standardized protocol. These include identification of an appropriate source of rabbits (e.g., an abattoir such as Pel-Freeze) to provide eyes to be used in the IRE, and inclusion of an explicit statement that that rabbits should not be bred and killed specifically for use in the IRE test method. The policies of the various U.S. regulatory agencies with respect to use of rabbits in the IRE that were used in previous tests or experiments needs to be reviewed and updated as it impacts the number of animals available for use in this test. The decision criteria used to identify ocular corrosives/severe irritants should be clearly identified and a rationale provided for how it was developed. For any future studies,

defined positive, negative, and benchmark substances need to be identified based on the proposed list of reference substances. In addition, the Panel proposed that the National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) facilitate the development of a standardized histopathology scoring system for corneal damage, along with an appropriate atlas with visual aids. In addition, the appropriate circumstances under which histopathology would be warranted should be more clearly defined. To maximize the likelihood of obtaining reproducible results, reference photographs for all subjective endpoints should be developed (e.g., corneal opacity, fluorescein penetration, histopathology) to aid training and transferability. A discussion of the use of proper safety precautions when handling animals and isolated eyes and awareness of the risk of contamination with potential zoonoses should also be included in the IRE BRD.

The Isolated Chicken Eye Test Method

The Panel concluded that the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) criteria for validation (ICCVAM 2003) have not been fully met for the ICE test method. Cited deficiencies include: the intralaboratory reliability of the ICE test method has not been adequately evaluated; the raw data from the three ICE studies included in this evaluation were not available for review; and detailed drawings/diagrams of the superfusion apparatus have not been made available to allow for transferability of the experimental setup. However, the Panel concluded that the ICE test method can be used in the identification of ocular corrosives/severe irritants in a tiered testing strategy, with specific limitations. Specifically, the Panel noted that alcohols tend to be overpredicted, while surfactants tend to be underpredicted. The Panel also recognized that solids and insoluble substances may be problematic in the ICE test method, since they may not come in adequate contact with the corneal surface, resulting in underprediction. Therefore, the Panel concluded that the low overall false positive rate (8% to 10%, depending on the regulatory classification scheme evaluated) indicates that the ICE test can be used at present to screen for severe eye irritants/corrosives. However, given the high false positive rates calculated for a small number of alcohols (50% [5/10]), the Panel noted that caution should be observed when evaluating ICE test results with this class of substances.

The Panel recognized that the recommended protocol is based on the original ICE protocol, which has changed only slightly since its development. However, there was concern expressed as to whether the appropriate number of eyes (n=3) is being used to ensure optimum performance. Therefore, the Panel recommended that the potential effects of using more than three eyes on the accuracy and reliability of the ICE test method be the subject of a formal study. The Panel also questioned the utility of using maximum mean scores, and thus to ensure optimum performance, recommended a formal evaluation of the most appropriate mathematical approach.

The Panel identified potential methodological areas of improvement to the protocol, including moving the superfusion apparatus to a horizontal position to obviate the need for test eye removal during dosing, adding centering lights to the optical pachymeter to ensure consistent central corneal thickness measurements across laboratories, and inclusion of concurrent negative and positive control eyes (at least 3 per group). In addition, histopathology, including

determining the nature and depth of corneal injury, was recommended for inclusion in the protocol when the standard ICE endpoints (i.e., corneal opacity, swelling, and fluorescein retention) produce borderline results. With this in mind, the development of a standardized scoring scheme using the formal language of pathology to describe any effects was advocated, along with defining the appropriate circumstances under which histopathology would be warranted. The Panel noted the need for reference photographs for all subjective endpoints (i.e., corneal opacity, fluorescein retention, and histopathology) to ensure consistency among laboratories.

Given the limited amount of ICE reliability data, additional studies using the recommended ICE test method protocol were suggested to better characterize the repeatability and the intra-and inter-laboratory reproducibility of the test method. The Panel recommended also optimization studies that were considered to be potentially useful for improving ICE test method performance. These studies included efforts to optimize the decision criteria used for identifying corrosives and severe irritants, an evaluation of the impact of routinely performing replicate experiments, and an evaluation of the impact of variations in the time between death and testing of the chicken eyes on test method performance.

The Panel specified that any optimization and validation studies should use existing animal data, if available, and that additional animal studies should only be conducted if important data gaps are identified. A minority opinion of one Panel member stated that no additional animals should be used for this purpose.

The Bovine Corneal Opacity and Permeability Test Method

The Panel concluded that the BCOP BRD proposed version of the test method has been shown to have adequate accuracy and reliability for detecting corrosive or severe eye irritants in the tiered testing scheme outlined in the BCOP BRD, with the following caveats:

- The test should not be used to identify corrosive or severely irritating ketones, alcohols, and solids. Further optimization and validation are necessary before these classes of materials can be assessed with this test.
- It needs to be confirmed that the BCOP test method can identify, as well as or better than the Draize test, those substances known to cause serious eye injury in humans. It appears from the list of chemicals tested that at least some of these substances have been tested in BCOP (e.g., floor strippers and heavy duty cleaners).
- A histopathological examination should be added to the test unless the test substance is from a class of materials known to be accurately predicted using only opacity and permeability in the BCOP assay.

The Panel concluded that the BRD proposed protocol for the BCOP test method is useful for identification of severe or corrosive ocular irritants in the tiered testing scheme outlined in the BCOP BRD, with the caveats noted above, as well as those noted below:

- 0.9% sodium chloride should be used instead of distilled water as the test substance diluent.
- Determination of osmolarity and pH of test solutions should be conducted.

- The optimum age range for cattle should be determined.
- Users should be aware of zoonoses, including the possibility of Bovine Spongiform Encephalopathy (BSE).
- Concurrent negative, positive, and benchmark controls should be used.

With respect to suggested modifications to improve performance (accuracy and reliability) of the recommended standardized protocol for the BCOP test method, the Panel recommended the following modifications:

- Use of the larger holder as suggested by Ubels et al. (2002, 2004).
- Re-examine the use of the calculated total score when the endpoint is severe injury only.
- Changes to the medium used to bathe the eyes, including a determination of whether fetal bovine serum is needed.

While the Panel believes these modifications are important, the Panel concluded that the data presented in the BCOP BRD support use of the BCOP assay in its current form for identifying ocular corrosives and severe irritants other than alcohols, ketones, and solids in a tiered testing strategy for regulatory hazard classification and labeling purposes.

The Panel also suggested that histopathological examination be added to the recommended test protocol unless the test substance is from a class of materials known to be accurately predicted using only opacity and permeability in the BCOP assay.

While actually a change to the BCOP method, the Panel suggested the possibility of using the porcine eye as a model for the human eye. The Panel recognizes that this change would require complete validation, but wants to be sure this possibility is considered for future work.

During a vote on Section 12.2 (Recommended Standardized Test Method Protocol) of the BCOP report at the Panel meeting, three panel members expressed minority opinions. Dr. Freeman abstained from voting on Section 12.2 because he believed the discussion on this section had not been satisfactorily resolved due to time constraints. Drs. Stephens and Theran did not agree with the final language presented for Section 12.2 because they believed the BCOP group members withdrew their original summary conclusion under undue pressure.

Regarding recommended optimization studies to improve performance (accuracy and reliability) of the recommended BCOP test method protocol, the Panel recommended using a larger holder similar to that suggested by Ubels et al. (2002), re-examining the use of the calculated total score when the endpoint is serious injury only, changing the medium used to bathe the eyes, using antibiotics if eyes are kept above 0 °C, and defining appropriate ages of donor animals. While the Panel feels these improvements are important, it believes the data presented in the BRD are sufficient for supporting use of the BCOP assay in identifying ocular corrosives and severe irritants, except for alcohols, ketones and solids, in a tiered testing strategy for regulatory hazard classification and labeling purposes.

With respect to the recommended validation studies to evaluate performance of the optimized BCOP test method protocol, the Panel concluded validation studies, or submission of additional

data supporting the three-minute exposure time suggested for volatile solvents, will be necessary before the BCOP test method can be recommended for use with alcohols and ketones. Validation studies or submission of additional data will be necessary before the BCOP test method is acceptable for solids. The Panel concluded the information in the BCOP BRD, along with the Panel's suggestions, is sufficient to support the use of this test method to identify severe irritants and corrosives, with the exception of alcohols, ketones and solids, in the tiered testing scheme described in the BRD.

The Panel concluded that an additional validation study is not necessary for the recommended additional histopathological examination to the BCOP test method. Although adding histology to the BCOP assay involves additional endpoints, current practice has not been to insist on validation of histopathological examination when it is added to an *in vivo* test method. A standardized histopathological scoring system was suggested by the Panel, but this should be arrived at by the experts in the field and will not require validation. NICEATM/ICCVAM should facilitate the development of a histopathological scoring system for corneal damage (with visual aids). Changes in the calculation method for the BCOP test score, or the use of the individual endpoint data instead of a calculated score also do not need to be validated.

When validation studies are conducted, the Panel believes the studies proposed in the BCOP BRD are appropriate but should be limited to the classes of test substances in question. Validation studies should be carefully planned. Tests should first be done to confirm that any modifications of the protocol do not decrease reliability. Once the inter- and intra-laboratory variability is defined, it will not be necessary to have a large number of laboratories test every chemical in the validation study. Validation should focus on the class of chemicals in question. The study should involve a very small number of experienced laboratories with only a limited number of duplicate samples at each laboratory.

Any validation or optimization studies should use existing animal data, if available. Additional animal studies should only be conducted if important data gaps are identified and such studies should be carefully designed to maximize the amount of pathophysiological information obtained (e.g., wound healing) and to minimize the number of animals used.

With respect to Section 12.3 of the BCOP report, one Panel member, Dr. Stephens expressed a minority opinion. The report leaves open the possibility of additional animal studies as part of this process. Dr. Stephens believes that no additional animal studies should be conducted for such optimization or validation exercises.

The Hen's Egg Test - Chorioallantoic Membrane Test Method

The Panel concluded that, for the purpose of detecting severe eye irritants in the tiered-testing strategy outlined in the HET-CAM BRD, the HET-CAM test has been shown to be useful for identification of severe or corrosive ocular irritants. The Panel stated that the high false positive rate was a limitation of the HET-CAM test method. It was proposed that positive results from the HET-CAM test method could be re-tested in a modified HET-CAM test method (e.g. using a lower concentration of test substance) to confirm the results. Alternatively, substances producing a positive result could be tested in a different *in vitro* test method (e.g., ICE, IRE,

BCOP). Substances producing negative results (e.g., HET-CAM score defined as nonirritant, mild irritant, or moderate irritant) would follow the tiered-testing strategy.

It was agreed that the most appropriate version of the HET-CAM test method for use in a tiered-testing strategy is the test method protocol recommended in the HET-CAM BRD. The proposed HET-CAM standardized test method protocol is adapted from the one by Spielmann and Liebsch (INVITTOX 1992). The proposed standardized test method protocol contains negative controls, solvent control (if appropriate), positive controls and benchmark controls (if appropriate). The method also recommends using the time required for an endpoint to develop as the criteria for assessing irritation potential (IS(B) analysis method). The Panel stated that procedures for applying and removing solids from the chorioallantoic membrane (CAM), which may adhere to the CAM and demolish the CAM upon removal, should be included in the standardized test method protocol provided in the HET-CAM BRD.

Due to the numerous variations in the test method protocols and different analysis methods that have evolved since the development of the test method, the Panel stated that the use of a standardized test method protocol in future studies would allow for new data to be generated. These data would allow further evaluation of the usefulness and limitations of the recommended test method protocol.

With regard to optimization of the recommended standardized test method protocol, the Panel stated that a retrospective analysis should be conducted to determine if different decision criteria might enhance the accuracy and/or reliability of the test method for the detection of ocular corrosives and severe irritants, as defined by the European Union (EU 2001), United Nations Globally Harmonized System (UN 2003), and the U.S. Environmental Protection Agency (EPA 1996) classification systems. The Panel proposed the use of a modular approach to validation to identify needed validation modules (e.g., interlaboratory reliability) and focus on evaluating those modules.

The Panel stated that the recommendation to optimize and to use an optimized method should not minimize the value of data already obtained with the method of Spielmann and Liebsch (INVITTOX 1992). As some laboratories already apply the method of Spielmann and Liebsch (INVITTOX 1992), the data generated in these laboratories should still be valid and be used for labeling of ocular corrosives and severe irritants. The Panel proposed that an optimized test method may be used when a positive finding is obtained in the HET-CAM test method of Spielmann and Liebsch (INVITTOX 1992); the substance could be re-tested in the optimized test method protocol.

The Panel further stated that inclusion of different endpoints (e.g., trypan blue absorption, antibody staining, membrane changes, etc.) for evaluation of irritancy potential may increase the accuracy of the HET-CAM test method. It was proposed that these additional endpoints may help reduce the number of false positives observed in the HET-CAM test. The Panel suggested that these endpoints could be included, but were not required, during optimization of the HET-CAM test method.

With respect to validation of the HET-CAM test method, the Panel agreed that if the test method were optimized and modifications made to the test method protocol had a major impact on the conduct of the study, a validation study should be conducted.

The Panel specified that any optimization and validation studies should use existing animal data, if available, and that additional animal studies should only be conducted if important data gaps are identified. A minority opinion of one Panel member stated that no additional animals should be used for this purpose.

The Panel further recommended that an evaluation be conducted to determine the relationship or predictability between the short-term effects observed in the HET-CAM and long-term effects observed in rabbits or humans be conducted. The Panel proposed that such an evaluation may provide additional support for the use of the HET-CAM method to assess the delayed and long-term effects of ocular corrosives and severe irritants.

Proposed List of Reference Substances for Optimization or Validation Studies and to Use in Establishing Performance Standards

The Panel reviewed the adequacy and completeness of the proposed list of reference substances and concluded that the list of proposed substances is comprehensive, the substances appear to be readily available and in acceptably pure form, and the range of possible ocular toxicity responses in terms of severity and types of lesions appears to be adequately represented. The Panel also concluded that, while it is recognized the selection of reference substances is in part limited by the availability of appropriate *in vivo* reference data, the current list has too many substances and is unwieldy, surfactants are over-represented and thus could be reduced in number, and more inorganic substances should be added, if feasible. The Panel also recommended that substances known to induce severe ocular lesions in humans should be included in the list, even in the absence of rabbit data. For all validation studies, Material Safety Data Sheets (MSDS) for the recommended substances should be provided (e.g., a coded MSDS); also prestudy safety briefings should be conducted routinely. Finally, the Panel recommended that an assessment based on the ranking of experimental data for severity for both the reference test method and the *in vitro* test, using the proposed reference substances, be conducted routinely.

For any future validation studies that are performed subsequent to protocol optimization, the Panel recommended that a two-staged approach be used to evaluate accuracy and reliability. Accordingly, the first stage would evaluate test method reliability using a subset of substances that could be tested in multiple laboratories, followed by a second stage encompassing a larger number of substances to evaluate test method accuracy. The Panel suggested that the accuracy assessment include a statistical analysis of the ranking of experimental data for severity for both the *in vivo* reference method and the *in vitro* test.

[This Page Intentionally Left Blank]